Relational Coherence and Persistent Rule-Following: The Impact of Targeting Coherence in a 'Non-Critical' Component of a Relational Network

Relational Coherence and Persistent Rule-Following: The Impact of Targeting Coherence in a 'Non-Critical' Component of a Relational Network

Ragnar Bern *[1], Tone Persdotter*[1], Colin Harte[2] and Dermot Barnes-Holmes[2, 3]

[1]Department of Psychology, University of Gothenburg, Gothenburg, Sweden

[2]Department of Experimental, Clinical and Health Psychology, Ghent University, Belgium

[3]School of Psychology, Ulster University, Coleraine, Northern Ireland, UK

Corresponding Author:

Colin Harte

Department of Experimental, Clinical, and Health Psychology

Ghent University

Henri Dunantlaan, 2

9000 Ghent

Belgium

Email: Colin.Harte@UGent.be

Authors' Note This article was prepared with the support of an Odysseus Group 1 grant awarded to the fourth author by the Flanders Science Foundation (FWO). Correspondence concerning this article should be sent to Colin.Harte@UGent.be

*****These authors contributed equally to this manuscript

Abstract

Rule-governed behavior has long been associated with generating insensitivity to direct contingencies of reinforcement. This insensitivity to environmental changes has also been implicated in human psychological suffering. Recent developments within Relational Frame Theory (RFT) have highlighted the importance of analyzing the dynamics of arbitrarily applicable relational responding (AARR) with regard to the impact of rules on human behavior. While previous research has focused on the impact of levels of derivation and coherence at the level of the relational frame, no published research to date has investigated the impact of coherence at the level of the relational network on rule persistence. Participants were first trained on a novel relational network that was either maximally coherent or partially incoherent before being exposed to a contingency switching Matching-to-Sample (MTS) task. Crucially, the current research aimed to investigate the impact of challenging the coherence of an aspect of the network that was not necessarily important for deriving the rule for responding on the MTS task. Results showed that coherence significantly impacted upon levels of rule resurgence, but no other measure of rule persistence. Correlational analyses indicated that manipulating coherence per se versus a control condition had a significant impact on specific self-report measures such as level of certainty. A post-hoc RFT interpretation of the findings is provided.

KEY WORDS: Rule-governed behavior; RFT; Coherence; Persistent rule-following

The importance of the impact of rules or instructions on human behavior has long been identified within the psychological literature (‘rules’ and ‘instructions’ will be used interchangeably throughout the current article). The concept itself, known as *rule-governed behavior,* was first introduced by B.F. Skinner (1966) within the context of an operant account of problem solving. At that time, rules were defined as contingency specifying stimuli that allowed the listener to problem solve without having to contact reinforcement contingencies directly. For example, a parent giving a child the simple rule “Look both ways before you cross the street to make sure no cars are coming” allows the child to learn important road safety skills without directly experiencing injury or worse from walking onto a busy road.

During the 1970’s and 1980’s, a plethora of experimental research emerged that focused on the impact of rules on human schedule performance. One of the key findings that emerged out of this work was that, for verbally-able humans, behavior under the control of instructions quite often led to what was termed an ‘insensitivity’ to direct contingencies of reinforcement (e.g., Catania, Shimoff, & Matthew, 1989). The term ‘insensitivity’ has been used to refer to the fact that verbal humans sometimes produce response patterns on schedules of reinforcement that differ from that of nonhuman animals (see Bentall, Lowem & Beasty, 1985, for example). The terms has also been used to refer to the related finding that providing instructions to human participants may produce insensitivity to changes in scheduled contingencies of reinforcement. For example, when instructed on how to earn reinforcers on a schedule of reinforcement, human participants tend to adapt less readily to un-cued changes in schedule contingencies relative to participants who were not initially instructed (see Hayes, 1989, for an early book-length review). This so-called rule-based insensitivity has since been widely argued as a potential moderating variable in human psychopathology (hereafter referred to as human psychological suffering; e.g., Baruch, Kanter, Busch, Richardson, & Barnes-Holmes, 2007; Rosenfarb, Newland, Brannon, & Howey, 1992; Zettle & Hayes, 1982).

In the behavior-analytic literature, rule-governed behavior and the insensitivity effect appeared to be a unique feature of human behavior. A second type of behavior that also appears to be unique in this regard is referred to as *derived relational responding*. This concept emerged with the seminal work of Sidman and colleagues (e.g., Sidman, 1971; Sidman & Tailby, 1982), the basic phenomenon of which came to be known as *stimulus equivalence* (see Sidman, 1994 for a book length treatment). The core finding was that after training a small number of relational responses (e.g., A=B and A=C), untrained and unreinforced responses often spontaneously emerged (e.g., B=C and C=B). Additionally, other untrained responses also often emerged when a specific function was trained to a stimulus participating in this newly derived relation (e.g., if A, B and C participate in an equivalence relation, and A is paired with a reinforcer, stimulus C may then acquire reinforcing functions in the absence of direct pairing). This latter effect has often been referred to as a derived transformation of functions. Crucially, while derived relational responding, including transformation of functions, appears to occur with relative ease in verbally-able humans, it is not readily observed in nonhuman animals or humans with severe language disabilities.

The extension of the early work on stimulus equivalence as a key explanatory tool for analyzing the complexities of human behavior came with the development of Relational Frame Theory (RFT), a behavior-analytic account of human language and cognition (Steele & Hayes, 1991; Hayes, Barnes-Holmes, & Roche, 2001). RFT suggests that stimulus equivalence should be considered as but one class of generalized operant behavior, and proposes that many others are possible. Specifically, these different operant patterns of derived relational responding are referred to as relational frames and include relations such as: similarity, difference, opposition, distinction, hierarchy, temporality, and deictics (see Hughes and Barnes-Holmes, 2016, for a recent extensive review). The generic concept of *arbitrarily applicable relational responding*

(AARR) is used to label these operant classes and their combination into increasingly complex relational networks.

While the study of derived stimulus relations and of rule-governed behavior have traditionally made little empirical connection, the conceptual link between the two has been quite strong. That is, both Sidman (1994) and Hayes et al. (2001; see also Hayes, 1989) argued that the human ability to engage in derived relational responding may be important for understanding how an instruction comes to specify contingencies of reinforcement. Indeed, some research has since suggested that derived relational responding could provide the basis for a technical analysis of rule-governed behavior, a suggestion which has been successfully modelled in the laboratory (O'Hora, Barnes-Holmes, Roche, & Smeets, 2004; O'Hora, Barnes-Holmes, & Stewart, 2014). Nevertheless, empirical work linking the two areas in any systematic way remains extremely limited. Recently, however, there has been a renewed effort to bridge the gap both conceptually and empirically between the two areas. This new line of research has involved conceptual developments within RFT itself, to which we now turn.

In recent years RFT has been somewhat 'updated' with the development of a new framework designed to help systematize RFT-based research more generally (see Barnes-Holmes, 2018; Barnes-Holmes, McEnteggart, & Barnes-Holmes, in press). This framework is known as the Hyper-Dimensional, Multi-Level (HDML) framework and provides a conceptual space for analyzing the dynamics involved in derived relational responding. The HDML conceptualizes AARR as varying along five levels and four dimensions. The five levels are based on conceptual and empirical analyses that have emerged from the literature on RFT (Hayes, et al., 2001) and are seen as increasingly advanced forms of relational development progressing from: (1) mutual entailment; (2) combinatorial entailment; (3) relational networks; (4) relating relations; and to (5) relating relational networks. We will not elaborate upon each of the levels here because they have been considered in many other sources since the publication of the seminal text on RFT (Hayes, et al., 2001; see Hughes & Barnes-Holmes, 2016, for a recent detailed summary). The HDML framework also divides the five levels along four dimensions: (1) coherence; (2) complexity; (3) derivation; and (4) flexibility.

A short description of the four dimensions is as follows. *Coherence* refers to the extent to which derived relational responding is generally predictable based on prior histories of reinforcement. For example, if you are told that 'X is larger than Y,' the derived response that 'Y is smaller than X' would be deemed coherent, but the response 'Y is the same size as X' would not (unless, of course, the wider context was modified to support an 'incoherent' response, such as 'Please respond to all questions with an incorrect answer'). *Complexity* refers to the level of detail or density of a particular pattern of derived relational responding. For example, the mutually entailed relation of coordination may be seen as less complex than the mutually entailed relation of comparison, because the former involves only one type of relation (e.g., if X is the same as Y, then Y is the same as X), but the latter involves two types of relations (if X is bigger than Y, then Y is smaller than X). *Derivation* refers to how well-practiced a particular instance of relational responding has become. Specifically, when a pattern of relational responding is derived for the first time, it is, by definition, highly derived (i.e., novel or emergent), and thus derivation reduces as that pattern becomes more practiced. Finally, *flexibility* refers to the extent to which a given instance of derived relational responding may be modified by current contextual variables. Imagine, for example, a young child who is asked to respond with the wrong answer to the question, "Which is bigger, a car or an bus?" The more rapidly the child responds with "car", the more flexible the relational responding (see O'Toole & Barnes-Holmes, 2009). Of course, flexibility is always context dependent and thus if the child had been "warned" previously not to give a wrong answer when asked to do so, it would be difficult to use the production of a correct or wrong answer as an indication of flexibility.

A detailed treatment of the HDML framework is beyond the scope of this article[1]. The critical point, however, is that when the study of derived relational responding is approached from within the framework, its potential to help researchers analyze the complexities and dynamics of human language and cognition (including most importantly in the current context, rule-governed behavior) may become apparent. As mentioned previously, a new line of research has since begun to bridge the gap between the work on persistent rule-following and derived stimulus relations. This research has sought to examine the impact of the different dimensions specified within the HDML on derived rule persistence (e.g., Harte et al., 2017, 2018, Harte et al., 2020). For example, a study by Harte et al. (2018) sought to investigate the extent to which the level of derivation involved within an experimentally derived rule impacted upon persistence in rule-following on a contingency-switching Matching-to-Sample (MTS) task. That is, would a rule that involved a novel derived relation, produce more or less rule persistence following an un-cued contingency reversal, when the relation involved high versus low levels of derivation. Across two experiments, participants were first trained on either novel mutually entailed relations (Experiment 1: A-B) or combinatorially entailed relations (Experiment 2: A-B/B-C) for either 1 block of training trials (high derivation) or 15 blocks (low derivation). Next, the novel mutually (i.e., A-B) or combinatorially (i.e., A-C) entailed relation was inserted into the rule required for correct responding on a subsequent MTS task. For the first 100 trials of the MTS procedure, the scheduled contingencies matched the derived rule, and participants gained points for correct responses. On the 101st trial, these contingencies reversed unbeknownst to participants. Thus, responding in accordance with the derived rule now resulted in a loss of points. Lower levels of

[1] The HDML framework was first described as *multi*-dimensional (see Barnes-Holmes, Barnes-Holmes, Luciano, & McEnteggart, 2017; Barnes-Holmes, Finn, McEnteggart, & Barnes-Holmes, 2018). The term 'hyper-dimensional' is now used to highlight a balanced emphasis on both entailment and transformation of functions, the properties of which define derived relational responding itself (see Barnes-Holmes, 2018; Barnes-Holmes, et al., in press).

derivation generated greater rule-persistence than higher levels, and this was the case for both mutually and combinatorially entailed relations.

In a subsequent study by Harte et al. (2020), the same basic paradigm was employed, but the coherence of the experimentally derived rule was manipulated through the use of performance feedback (it was assumed that providing feedback for “correct” derived responding would likely increase coherence). Across two experiments, participants were first trained on the same baseline relations as Harte et al. (2018). In Experiment 1, participants were then retrained on the same baseline relations for two further blocks of trials, with one group receiving feedback on their performances and another group receiving no performance feedback. Following baseline training in Experiment 2 however, participants were directly tested on the derived A-C relations for two further blocks of trials. Once again, half of the participants received feedback on their performance while the other half did not. Participants in both experiments then completed the contingency-switching MTS task. Feedback, differentially impacted upon persistent rule-following, but only when participants were given the opportunity to derive the A-C relations. Specifically, in Experiment 2, feedback appeared to increase at least on type of rule persistence.

The primary purpose of the current study was to extend and elaborate the research conducted thus far on the impact of coherence on persistent derived rule following, but at the level of the relational network. That is, would a condition involving a relational network that contained some feature of incoherence produce more or less persistence in rule-following on the same contingency switching MTS task than a condition involving a relational network that was maximally coherent? Specifically, the current study involved training participants on a six member relational network that was maximally coherent in one condition (i.e., train A=B=C=D=E=F; *reinforce* the derived F=D relation) but contained an element of incoherence in another condition (i.e., train A=B=C=D=E=F; *punish* the derived F=D relation). Crucially, we deliberately did *not* seek to undermine the coherence of the part of the network that would be involved in deriving the rule for completing the subsequent MTS task (which was restricted to the A=B=C part of the network). Instead, coherence was manipulated among the D=E=F members of the network. Given the highly exploratory nature of the current study, we refrained from making any formal predictions.

Before proceeding, it is worth noting that a reviewer of an earlier version of the current article sought clarification concerning why we refrained from making any formal predictions in light of previously published studies, and the extent to which they articulated with the HDML (previously the MDML) framework. The reason for not making any formal predictions is that the current study is unique in that the coherence in the relational network was manipulated in a part of the network which was then *not* directly involved in the derived rule. Given that no previous study, published or unpublished, to our knowledge, had attempted to manipulate coherence in a non-critical part of the network, there appeared to be no solid basis upon which to make a specific prediction. The research, therefore, is better characterized as being based on a "what would happen if..." strategy rather than a hypothetico-deductive approach in which previous research, combined with a formal theoretical model, is used to predict X (see Chiesa, 1994, who argues that radical behaviorism is better characterized as involving the former over the latter approach).

It is also worth emphasizing that the HDML should *not* be seen as providing the basis for making formal predictions, or more specifically, formal hypotheses. As noted by Barnes-Holmes et al. (2017), ". . . the MDML [now HDML] . . . is not a new model that makes specific predictions. Rather, the MDML is a framework that seeks to make explicit what basic researchers in RFT have been doing implicitly since the theory was first subjected to experimental analysis. In this sense, the MDML may be seen, in part, as a framework for orienting basic researchers in RFT to new possibilities for future research." (p. 435). Indeed, it is also worth noting that the HDML may be seen as supporting a functional-analytic abstractive approach to science that differentiates behavior analysis from the hypothetico-deductive strategy that characterizes mainstream psychology. For example, even in a situation in which a behavior-analytic researcher fails to replicate a previous finding, the HDML serves to highlight the complex and dynamical variables that may be at play. As such, even a very small difference in just one of the many variables involved could lead to a different outcome than the one obtained in the earlier study. Indeed, if a counter-intuitive result emerges in a study the HDML may prove useful in orienting the researcher towards the likely variable or variables that could explain the result in question. The HDML may be seen, therefore, as an example of the sine qua non of the behavior-analytic inductive approach to science. We shall return to this issue in the Discussion.

Method

Participants

216 undergraduate and graduate students (138 females, 70 males, 5 other, and 3 preferred not to answer) were recruited through random convenience sampling at the University of Gothenburg, Sweden. Their ages ranged from 18-57 years (*Mean range* = 22-25) and 94.9% spoke Swedish as their first language. Participants were randomly assigned to one of three conditions: High Coherence, Low Coherence, or Control. The Control Condition was further sub-divided into two conditions for counter-balancing purposes (described subsequently): Control-Faster and Control-Slower. The data from 75 participants (31 from High Coherence; 29 from Low Coherence; and 15 Control) were excluded because they failed to meet a number of specific task performance criteria (described subsequently), leaving N = 141 for analysis (44 in the High Coherence; 47 in the Low Coherence; 23 in the Control-Faster; and 27 in the Control-Slower).

Setting

The experiment was conducted in the computer laboratories at Gothenburg University. Between one and 20 participants were present in the computer room completing the experiment at any given time (i.e., a 'cafeteria style' setting in which participants were free to turn up to the laboratory to complete the study within specified time windows). Participants were always placed as far away from each other as possible and were instructed by the experimenters before entering that they must remain silent for the entirety of the experiment. Participants were shown to a computer and desk upon which a written instruction reminded them not to speak with one another and to start the experiment whenever they were ready. All other instructions were provided on the computer screen. The experimenters remained present for the entirety of the experiment.

Apparatus and Materials

The experiment involved one self-report measure (Depression Anxiety and Stress Scales [DASS]-21; Lovibond & Lovibond, 1995) and two computer-based tasks made in Qualtrics (a Coherence Task and MTS task). Participants completed all aspects of the experiment on a standard Dell personal computer. The instructions and stimuli involved were all presented in Swedish, but their English translations are presented here.

Coherence Task. The Coherence Task consisted of eight individual trials comprised of six task-relevant trials and two task-irrelevant trials. Each individual trial was composed of (i) between two and three short statements, (ii) a question about these statements, and (iii) between two and three response options.

The first three task-relevant trials involved the first part of a six-member network (A=B=C), while the last three task-relevant trials involved the latter part of this network (C=D=E=F; see Table 1 for an illustration of these complete networks per condition). The statements, question, and response options that comprised the task-relevant trials in the first part of the network (A=B=C) are presented in Figure 1 (left-hand side). In the first statement, the word “KROS” (C) was coordinated with the word “ZID” (B), and “ZID” was then coordinated with “LEAST LIKE” (A). Hence, participants could derive that “KROS” had the same meaning as “LEAST LIKE”. Participants could select a response from the options “LEAST LIKE”, “MOST LIKE” or “SAME”, when asked “What does KROS mean?”. Correct responding involved choosing the “LEAST LIKE” response option. This trial was considered task-relevant because it enabled participants to derive that the nonsense word “KROS” had the same meaning as the phrase “LEAST LIKE”, the meaning of which would be necessary to accurately interpret the rule for responding in the subsequent MTS task.

INSERT TABLE 1 HERE

Table 1 An illustration of the relational networks trained per experimental group

INSERT FIGURE 1 HERE

Fig 1 Illustrations of the task-relevant trials presented to the High and Low Coherence Conditions in the Coherence Task. These were similar for the Control Conditions except that “LEAST LIKE” was replaced with “FASTER THAN” or “SLOWER THAN” (left-hand side) depending on whether participants were in the Control-Faster or Control-Slower Condition. As such, the response options were also altered to reflect this change. The trial presented on the right-hand side was the same for each condition

The final three task-relevant trials were comprised of the last four stimuli of the network (i.e. C=D=E=F). The statements, question, and response options that comprised the task-relevant trials in the first part of the network are presented in Figure 1 (right-hand side). In the first statement presented on each of these three trials, the word “KROS” (C) was coordinated with the word “VEK” (D), “VEK” was then coordinated with “JUM” (E), and finally “JUM” was coordinated with “POM (F)”. Hence, participants could derive that “VEK” was coordinated with “POM". Participants could select a response from the options “YES” and “NO” and correct responding involved selecting the “YES” response option, when asked “Are VEK and POM the same?” This trial was again denoted as task-relevant because it contained stimuli that participated in a network in which all of the stimuli could be derived as “Least Like”, although derivation between the D, E, and F stimuli and “Least Like” was never tested.

The two task-irrelevant trials are presented in Figure 2. In the first task-irrelevant trial (see Figure 2 left-hand side), “SAM” was said to be younger than “TOM”, and “TOM” younger than “PAT”. Participants could select a response from the options “TOM” “SAM”, and “PAT” and correct responding involved selecting “PAT” when asked “Who is the oldest?” The second task-irrelevant trial was similar, except that the relations between the three stimuli varied along the dimension of strength instead of age (see Figure 2 right-hand side). These trials were denoted as task-irrelevant because nothing derived from them could be used to inform responding on the subsequent MTS task. These trials were included so that participants were required to derive a number of relations across the tasks rather than simply learning to pick the same comparison (e.g., “Least Like”) on every trial.

INSERT FIGURE 2 HERE

Fig 2 An illustration of the task-irrelevant trials presented to all conditions in the Coherence Task

Each individual trial was followed by a 7-point Likert scale in which participants were asked to rate how certain they were about their answer to the question on that trial. Responses ranged from 1 (very uncertain) to 7 (very certain). Immediately following the fourth trial and associated Likert scale, participants were presented with an open-text question in which they were asked what “KROS” meant, to which they manually typed their response.

The foregoing paragraphs described the Coherence Task for the High and Low Coherence Conditions. This task was similar for the Control Conditions, except that the word “LEAST LIKE” was replaced with “SLOWER THAN” or “FASTER THAN” within the task relevant trial-types. All trial-types in the Control Condition were thus irrelevant to completion of the subsequent MTS task.

MTS Task. This task comprised of 150 trials, during each of which a sample stimulus (random shape) was presented at the top of the screen, accompanied by three comparison stimuli (all random shapes but none identical to each other nor the sample) along the bottom of the screen (see Figure 3). Each of the comparison stimuli always varied with respect to its similarity to the sample stimulus in the following ways: One comparison was evidently the most similar to the sample in that it was the same base shape but contained minor variations (see center Figure 3); one comparison was also evidently similar to the sample (and to the previous comparison stimulus), but its shape contained more variations (see left-hand side Figure 3); the final comparison was evidently the most different from the sample, comprised of a different shape of little or no common features (see right-hand side Figure 3). An individual stimulus set was comprised of one sample stimulus and three-comparison stimuli combinations. The task was made up of a total of 54 stimulus sets, each set being presented at least once and no more than three times across the 150 trials. To respond, participants used the mouse to select the comparison stimulus that they wished to choose.

INSERT FIGURE 3 HERE

Fig 3 An example of a single trial and single stimulus set presented in the MTS task

Questionnaire. The DASS-21 (Lovibond and Lovibond, 1995) is comprised of three sub-scales measuring depression, anxiety and stress with 7-items per sub-scale across a total of 21 statements (e.g., an item from the depression subscale is “I felt that I had nothing to look forward to”). All items are rated with respect to participants’ experiences within the last week on a scale from 0 (*Did not apply to me at all*) to 3 (*Applied to me very much or most of the time*). Sub scales are scored independently and indicate normal, mild, moderate, severe, and extremely severe levels. Higher scores indicate greater levels of psychological distress. The measure has demonstrated excellent internal consistency: depression ($\alpha = 0.88$); anxiety ($\alpha = 0.82$); stress ($\alpha =$

0.90); and total DASS (α = 0.93). The Swedish version of the scale was employed in the current experiment which has yielded similar levels of internal consistency (Alfonsson, Wallin, & Maathz, 2017).

Procedure

The experiment was comprised of three stages: 1. the Coherence Task; 2. the MTS task; and 3. the DASS-21 questionnaire, always conducted in that order. Before beginning, participants were told "You will now begin the first part of the experiment. What you learn in the first part might be of help in the second part. You will sometimes receive feedback on your answers and sometimes not."

Coherence Task. Before beginning, the following instruction was presented on the screen to participants: "You now begin part 1 of 2. What you learn in the first part might be of help in the second part. You will sometimes receive feedback and sometimes not." During this task, participants received eight trials (six task-relevant and two task-irrelevant) and one open-text question. Each trial was followed by a Likert scale that ranged from 1 (very uncertain) to 7 (very certain) on which participants rated how certain they were of their answer. This task sequence was as follows: 1. task relevant trial-type and Likert certainty scale; 2. task irrelevant trial-type and Likert certainty scale; 3. task relevant trial-type and Likert certainty scale; 4. task irrelevant trial-type and Likert certainty scale; 5. open-text question; 6. task relevant trial-type and Likert certainty scale; 7-9. task relevant trial-types and Likert certainty scales. After the first two trials, all participants received a reminder that "What you learn in this part might be relevant in the later part of the experiment". After the fourth trial, an open text box was presented and asked participants to type the meaning of the word "KROS" based on what they had just learned.

Performance feedback was provided on the final three trials, which asked participants if VEK was the same as POM. The nature of this feedback was manipulated between conditions. For participants in the High Coherence Condition, providing the correct response on these trials resulted in the feedback “correct answer” appearing on the screen in green, while an incorrect response on these trials resulted in the feedback “incorrect answer” appearing on the screen in red. For participants in the Low Coherence Condition, responding always produced the feedback “incorrect answer”, regardless of whether participants made a correct or incorrect response. The purpose of the differential performance feedback was to reduce the level of coherence in the network, thus making it a Low Coherence Condition (i.e., C=D=E=F≠D). Note that the coherence of the first part of the network (KROS=”LEAST LIKE”) was not “challenged”, and thus any difference between the High and Low Coherence conditions could not be attributed to a difference in coherence within the first part of the network.

For participants in the Control Condition, the Coherence Task was similar to that employed in the High and Low Coherence Conditions but for one key difference: “LEAST LIKE” was replaced with either “SLOWER THAN” (for the Control Slower Condition) or “FASTER THAN” (for the Control Faster Condition). The Control Condition was subdivided into ‘Faster’ and ‘Slower’ to control for the possibility that KROS may have been related more readily to “faster” or “slower” based on some unforeseen controlling variable (see Harte et al., 2018 for a discussion regarding cross-modal correspondence). All trials were task-irrelevant for this condition, given that deriving that “KROS” meant “Faster” or “Slower” would not aid participants in successfully completing the subsequent MTS task.

MTS Task. Upon beginning this part of the experiment, the following instructions appeared on the screen “You will now begin part 2 of 2. Try to gain as many points as possible. Your total score will be shown at the end of the experiment. Remember - what you learned in the first part might be of help to find the right answer in the next part.” At the bottom of the screen, the text “I understand” was next to a box which participants could click with their cursor when they understood, followed by a separate button indicating that they were ready to proceed. On the next screen, all participants were presented with an example of a screen in the task and what they were being asked to do. The word “Example” was printed at the top of the screen accompanied by a sample stimulus set in the center of the screen. That is, a sample stimulus was shown with three target stimuli underneath, as would be the case for the entire task. The following text was presented underneath this example: “Look at the upper symbol. Respond by choosing the lower symbol that is KROS the upper symbol.” Once again, the text “I understand” was presented on the screen next to a box which participants could click with their cursor when they understood, followed by a separate button indicating that they were ready to proceed. Upon proceeding, participants began the main task. During the first 100 trials of the 150 trial MTS task, participants were required to select the comparison stimulus that was least like the sample, and received a point for doing so. On the 101st trial, the task contingencies were reversed, un-cued to participants. Correct responding now involved selecting the comparison stimulus that was most like the sample rather than least like. If a correct response was emitted, the message “CORRECT +1 You earned one point” was presented in green lettering. If an incorrect response was emitted, the message “INCORRECT -1 You lost one point” was presented in red lettering. While participants’ total number of points accrued for data analysis, this was not presented on the screen while responding throughout the task (the total only appeared when the task was completed).

Questionnaires. After the MTS task, participants completed the DASS-21.

Results

Exclusions

Before the main analyses, a number of strict performance inclusion criteria were applied to the current data set. First, in order to exclude anyone who did not successfully make the required derivation in the Coherence Task, participants who gave an inaccurate (i.e., “MOST LIKE”) response to the fifth trial (i.e., “What does KROS mean?”), and an inaccurate (“NO”) response to the last trial (i.e. “KROS is the same as VEK. VEK is the same as JUM. JUM is the same as POM. Is VEK the same as POM?”) were excluded from the analysis. The former criterion resulted in the loss of 7 participants (3 from High Coherence, 3 from Low Coherence, and 1 from Control-Slower), while the latter resulted in the loss of 5 participants (1 from High Coherence, 2 from Low Coherence, 1 from Control-Faster and 1 from Control-Slower; N = 204 remaining).

Second, the strict performance criteria employed in Harte et al. (2017, 2018) were applied to the current data. That is, participants in the High and Low Coherence Conditions were required to respond correctly on at least eight out of the first ten trials on the MTS task to be included in the analysis. This criterion aimed to reduce the likelihood that participants learned to match based purely on trial and error, and instead were able to apply the rule they had derived from the first part of the experiment (the Coherence Task). The data from participants who did not meet this accuracy criterion in the initial trials were excluded. This resulted in the loss of 35 participants (16 from High Coherence and 19 from Low Coherence; N = 169 remaining). This accuracy criterion was not applied to the Control Conditions since it was expected that few participants would meet it (i.e. they had no rule to follow during their initial exposure to the MTS task). Nevertheless, 28.3 % of Control participants emitted at least 8/10 correct responses in the first 10 trials.

Additionally, participants from all three conditions were required to achieve 80 correct responses out of the first 100 on the MTS task (before the switch in task contingencies). This criterion was based on the assumption that an acceptable number of control participants would have adapted to the contingencies across the first 20 trials (Harte et al., 2017). The data from participants who did not meet this accuracy criterion were excluded from the analyses. This resulted in the loss of 28 participants (11 from High Coherence, 5 from Low Coherence, 8 from Control-Faster, and 4 from Control-Slower). The application of the foregoing performance criteria left N = 141 for analysis.

Preliminary Analyses: Control Conditions

Before conducting the primary analyses, possible differences between the two Control Conditions as a result of the "Faster Than/Slower Than" manipulation was assessed. First, two independent *t*-tests were conducted to compare the mean scores on all measures of rule-persistence employed in the current experiment (described subsequently) to each other. No significant differences in mean scores were found between the Control-Faster and Control-Slower Conditions, all *p*s > .05. Second, three independent *t*-tests were conducted to compare participants' mean DASS Depression, Anxiety, and Stress scores between the Control-Slower and Control-Faster Conditions. Again, no significant differences in mean scores were found, all *p*s > .05. Thus, the Control-Faster and Control-Slower Conditions were collapsed and analyzed as one single Control Condition.

Certainty Analyses

In order to determine if the primary coherence manipulation resulted in different levels of certainty in the experimentally trained networks, the means of each conditions' certainty scores (as measured by the final three Likert scales) in the last three trial-types in the Coherence Task (i.e. in which participants were given different feedback depending on which condition they were in) were calculated. A one-way between-participant analysis of variance (ANOVA) was conducted with condition (High vs. Low vs. Control) as the independent variable, and Coherence Task mean certainty score as the dependent variable. A significant difference among the three groups emerged, $F(2, 134) = 14.72$, $p < .001$, $\eta 2 = 0.17$. Post-hoc comparisons using Bonferroni corrections indicated that participants in the Low Coherence Condition ($M = 5.43$, $SD = 1.25$) reported significantly lower levels of certainty than both the High Coherence ($M = 6.36$, $SD = .94$) and the Control Condition ($M = 6.43$, $SD = .83$), *p*s < .001. No significant difference was found between the High Coherence and Control Condition.

Rule-Persistence Analyses

In order to analyze any differences in rule-persistence between conditions, the data from the final 50 MTS trials presented after the contingency switch were measured in three separate but related ways (consistent with Harte et al., 2018). These measures are referred to as Rule Compliance, Contingency Sensitivity, and Rule Resurgence.

The first measure, Rule Compliance, was defined as the total number of responses (out of 50) that were consistent with the original rule (choose the image that is "KROS" [meaning least like]), but inconsistent with the reversed task contingencies. Figure 4 (left-hand side) presents the mean Rule Compliance scores and shows differential scores across conditions. That is, participants in the High Coherence Condition emitted a larger mean number of responses (M = 22.72, SD = 20.65) in accordance with the original rule than did the Low Coherence (M = 19.09, SD = 18.50), and Control (M = 5.94, SD = 7.00) conditions. A one-way between subjects ANOVA was conducted and confirmed a significant difference in Rule Compliance scores among the three groups, $F(2, 134) = 14.11$, $p < .001$ $\eta2 = 0.21$. Planned comparisons using Tukey's HSD indicated that the mean Rule Compliance score for the Control Condition was significantly lower than that of both the High Coherence and Low Coherence Conditions (ps < .001). However, there was no significant difference in scores between High and Low Coherence (p = .55).

INSERT FIGURE 4 HERE

Fig 4 Mean rule compliance scores (left panel) and contingency sensitivity scores (right panel) with standard error bars for the Control, High Coherence, Low Coherence Conditions

Contingency Sensitivity was defined as the point at which participants stopped responding in accordance with the initial rule, and began responding in accordance with the reversed contingencies. That is, consistent with Harte et al. (2018), contingency sensitivity was defined as a pattern of at least three consecutive responses that were not in line with the original instruction, with at least one of these responses in line with the reversed contingency. Therefore, in principle, a participant could stop following the instruction and instead choose the stimulus that resulted in a loss of points (i.e., the stimulus that was ‘mid-way’ between the most-like and least-like images), but could only do this for two out of the three responses. This requirement ensured that the term ‘contingency sensitivity’ was appropriate, given that a participant must obtain at least one point when they stopped following the original rule. In any case, a post-hoc analysis of the individual participant data indicated that all participants chose the most-like comparison across all three responses (gaining three points), hence demonstrating contingency sensitivity.

Figure 4 (right-hand side) presents participants mean Contingency Sensitivity scores and once again shows differential mean scores across conditions. That is, participants in the High Coherence Condition took longer to exhibit contingency sensitive responding (M = 23.05, SD = 19.49) than did the Low Coherence (M = 18.50, SD = 18.23) and Control (M = 7.86, SD = 6.67) conditions. A one-way between subjects ANOVA was conducted and confirmed a significant difference in Contingency Sensitivity scores among the three groups, $F(2, 134) = 11.83$, $p < .001$ $\eta2 = 0.18$. Planned comparisons using Tukey's HSD indicated that the mean Contingency Sensitivity score for the Control Condition was significantly lower than that of both the High Coherence ($p < .001$) and Low Coherence conditions ($p = .003$). However, there was again no significant difference in scores between High and Low Coherence ($p = .36$).

Rule resurgence attempted to measure responding that was consistent with the initial rule (via percentage or responses), but occurred after a participant had emitted 3 consecutive responses in accordance with the reversed contingencies (hence the term *resurgence*). Figure 5 presents the density and range of participant data in each condition. Moderate levels of resurgence were observed in both Low Coherence and Control conditions, with a greater range of resurgence scores observed in the Low Coherence condition, and notably reduced levels of resurgence in the High Coherence Condition. Given that the data were severely skewed, a

Kruskal Wallis test was employed to explore any differences between the three conditions. The analysis proved to be marginally significant (High Coherence *Md* = 0.00, Low Coherence *Md* = 2.24, Control Condition *Md* = 2.22, $h(2) = 5.58$, $p = .06$). Planned comparisons using Mann Whitney U-tests indicated that the High and Low Coherence Conditions differed significantly from each other ($p = .02$). No statistically significant differences emerged between the High or Low Coherence Conditions and the Control Condition (*p*'s > .18).

INSERT FIGURE 5 HERE

Fig 5 Box plots with violin element illustrating the density and distribution of participant rule resurgence scores for the Control, High Coherence, and Low Coherence Conditions

Correlational Analyses

Given the differences recorded for both rule compliance and contingency sensitivity between the Coherence Conditions and the Control Condition, but not between the two Coherence Conditions themselves, correlational analyses were conducted separately for the Control Condition but together for both Coherence Conditions. In the Control Condition, both rule compliance ($r = .35$ $p = .01$) and contingency sensitivity ($r = .33$, $p = .02$) correlated positively with DASS stress, suggesting that more persistence with the original instruction predicted higher stress. Relatedly, a positive correlation was also found between contingency sensitivity and anxiety on the DASS ($r = .33$, $p = .02$), suggesting that greater rule persistence predicted higher levels of anxiety. As a result of the significant group differences recorded on the rule resurgence measure, separate correlational analyses were conducted for each condition between resurgence and the self-report scales, however no significant correlations were found (all *p*s > .11)

Discussion

The current study sought to extend research exploring the behavioral dynamics involved in persistent derived rule following, with a particular focus on the dimension of coherence, as specified within the HDML framework. Coherence was manipulated by reinforcing versus punishing an aspect of a novel relational network (i.e., the F=D derivation of the A=B=C=D=E=F network) on which participants were experimentally trained. Crucially, the part of the network necessary for deriving the rule for responding on the subsequent MTS task (i.e., the A=C derivation within the same network) was *not* targeted. In the Low Coherence condition, correctly deriving the F=D relation was punished, while in the High Coherence condition this relation was reinforced. Thus, one group derived a rule for responding on the subsequent MTS task that participated in a maximally coherent network (High Coherence Condition), while the other derived a rule that participated in a network that was relatively less coherent (Low Coherence Condition). Results indicated that coherence impacted rule resurgence, but not rule compliance or contingency sensitivity. That is, the Low Coherence group resurged back to the original rule for significantly more responses than did the High Coherence group. It should be noted that an effect for rule resurgence was also found in the previous Harte et al. (2020) study, which also manipulated coherence. In the previous study, however, the researchers manipulated coherence through the provision versus non-provision of performance feedback at the level of the relational frame, and also targeted the coherence of the derived rule directly.

The current study also found that certainty about the derived meaning of the rule differed significantly among the three conditions. That is, participants in the High Coherence condition reported significantly higher levels of certainty for the meaning of KROS (Least like) than did participants in the Low Coherence group. Crucially, the derived relation (A=C) was not punished in the Low Coherence group directly (only the D=F derivation was targeted). Thus, it appears that the impact of ‘challenging’ the coherence of part of a relational network may propagate through the entire network, at least in terms of self-reported certainty pertaining to the ‘meaning’ of a specific member of the network.

As noted above, an effect for rule resurgence was observed in a previous study of a similar nature (Harte et al., 2020). In that study, coherence was manipulated through the presence versus absence of performance feedback for the derived relation that was actually contained within the rule. As such, the impact of coherence on rule resurgence (and perhaps self-reported certainty) seems worthy of further investigation. At the present time it remains unclear why the coherence manipulation appeared to impact upon rule resurgence and self-reported certainty, but not rule compliance or contingency sensitivity. We are still relatively early in a research program that has sought to examine the role of the variables highlighted within the HDML framework on the impact of relational networks on persistent rule-following. An earlier study (Harte et al., 2018) indicated that levels of derivation, as opposed to coherence, did impact rule compliance and contingency sensitivity, but not rule resurgence. It may be, therefore, that the different properties of derived relational networks identified within the HDML framework may impact upon different measures of persistent rule-following. It seems premature to speculate exactly why this might be the case, but it is certainly worth noting here because it could provide the basis for potentially important future experimental analyses.

Although it was not a key focus of the current research, it is interesting that significant correlations were obtained between persistent rule following and self-reported levels of stress and anxiety, but only in the Control condition. Such a finding appears to be broadly consistent with the long-established claim that excessive rule-following is involved in, or may characterize, human psychological suffering (e.g., Zettle & Hayes, 1982; Hayes, Strosahl, & Wilson, 1999). In making this argument, it should be noted that the control condition likely involved some element of rule-following, given that participants were asked to earn as many points as possible. Thus when the contingencies changed during the MTS task, an inherent conflict was created between the rule provided by the experimenter and a self-generated rule that may have emerged during exposure to the original task contingencies. This finding could be particularly important given that the evidence base for the relationship between rule-following and psychological suffering remains extremely limited. In any case, the general experimental preparation employed in the current work (i.e., with the control group) could prove useful in pursuing this issue.[2]

One limitation of the current study is the considerably high level of attrition observed, particularly in the High and Low Coherence conditions. While it remains unclear at the current time exactly why the level of attrition was quite so high, it is interesting to note that similarly high levels of attrition were observed in a study by Harte et al. (2017), in which a similarly constructed training task was employed. Interestingly, in follow-up studies by Harte et al. (2018; under review), the derivation task used to train the novel relations was replaced with a training procedure based on the implicit relational assessment procedure (IRAP), and a dramatic reduction in levels of attrition was observed. Perhaps the specific format of the training tasks used in the current study and that of Harte et al. (2017) is itself part of the problem. Thus, it may be advisable to use the Training IRAP in future studies of a similar nature in the place of the type of task employed here. In any case, the levels of attrition observed within both coherence conditions in the current study were more or less equal (i.e., 31 participants from High Coherence and 29 participants from Low Coherence). It seems unlikely, therefore, that the differences observed between these two groups were a result of attrition alone.

[2] Harte, et al (2017) reported a negative correlation between self-reported stress (on the DASS) and rule-persistence in their direct rule condition, whereas a positive correlation was obtained in the control condition in the current study. A direct comparison across the two studies is unwise, however, because they differed in many respects. For example, the size of the networks differed, and the procedural details of the tasks that trained and tested these networks also differed in a number of respects; in addition the type of feedback that was presented during the MTS tasks also differed (see below). At the present time therefore it is difficult to explain why these differences in correlations between stress and rule-persistence emerged across the two studies, without becoming unreasonably speculative. In this context, it is worth noting that future studies might employ a measure of stress that directly targets current state rather than the wider reporting window associated with the DASS (i.e., emotions experienced during the previous week).

Another potential limitation of the current study is in the fact that the MTS task did not tally total points as participants progressed through the task. While highly speculative, if this feature was included in the procedure, perhaps somewhat more pronounced effects would have been observed. Indeed, studies using a similar preparation in the investigation of rule persistence that have found more pronounced effects have typically employed a cumulative record of points that participants could see during the task (i.e., Harte et al 2017, 2018, 2020). Perhaps, seeing points accrue and/or decrease from a total allocation, impacted upon the conflict between choosing to persist with a now 'incorrect' rule or continuing to earn points following the contingency reversal. Indeed, it is this very conflict that makes rule persistence in the face of competing reinforcement contingencies so interesting. Future research could attempt to replicate the current study, but ensure that a visible cumulative record of points is included in the procedure.

A related issue pertains to how we conceptualize the type of feedback that was presented in the current, and indeed related, studies. It may be tempting to view feedback as simply reinforcing versus punishing stimuli. For verbally-able humans, however, it may be more accurate to conceptualize feedback as participating in a derived relation. For example, when the feedback consists of a message indicating that a point has simply been added or subtracted as in the current study, the relation between the two forms of feedback could be seen as participating in a frame of opposition. In contrast, when the feedback involves an increase or decrease in an ongoing tally of total points gained or lost, the feedback may be conceptualized as participating in a frame of comparison. In the latter case, gaining and losing points may have increasing or decreasing appetitive/aversive functions, respectively. Or more informally, the feedback may be more powerful when participants directly witness their tallied points actually growing or diminishing on screen. The impact of these two types of feedback, one that may be defined as oppositional versus one that seems to be more comparative, requires systematic analysis in its own right and certainly needs to be taken on board when exploring rule-following in the face of reversed feedback contingencies.

One final interesting issue that arises from the current study is that resurgence in rule following was observed in the condition deemed to be *low* in coherence. In contrast, in Harte et al., (2020), the condition deemed to be *high* in coherence (i.e., Feedback Condition) produced greater resurgence in rule following. One possible explanation is that attempting to reduce coherence by punishing coherent relational responding with regard to a non-critical part of the network may have reduced the coherence properties of feedback itself. Thus, when feedback was used to punish rule-consistent responding on the MTS task following the contingency reversal, it was less effective in suppressing rule following in the Low Coherence group, and therefore resurgence became more likely in this group. More informally, the feedback presented during the MTS task was undermined somewhat because it had proved to be unreliable in an earlier part of the study when it was used to punish coherent relational responding. Although this interpretation is post-hoc, it does serve to highlight why the HDML may be useful in orienting the researcher towards analytic-abstractive concepts such as coherence. In this case, the framework has been used to speculate about the potential change in the reliability of the feedback itself when the feedback did not cohere with earlier derived relational responding. As noted in the introduction, this general approach to using the HDML framework, as an instrument for inductive behavior-analytic research, is to be contrasted with using it as a theoretical model for making very specific hypothetico-deductive predictions. In this sense, of course, the HDML framework cannot be falsified in a hypothetico-deductive manner, but the extent to which it continues to assist researchers in identifying potentially important variables in the functional-analytic abstractive approach to behavioral science will ultimately determine its long-term survival.

In closing, the current work again highlights the complexities involved in attempting to examine the extent to which rule-governed behavior accounts for so-called insensitivity to contingencies of reinforcement. Intuitively, the basic idea clearly has some face validity, and indeed has been widely used in the literature in clinical behavior analysis to argue that excessive rule-following may be implicated in human psychological suffering. On balance, the empirical evidence in this area remains sparse and in some cases contradictory. The current findings, and those reported in similar recent studies indicate that the we need a far more sophisticated analysis of rule-governed behavior itself, particularly in terms of the dynamics involved in the relational networks that serve as instructions or rules. Without such analyses, genuine progress in this important area of research will likely not emerge.

Declarations and Compliance with Ethical Standards

Conflict of Interest: The authors declare they have no conflict of interest

Funding: This article was prepared with the support of an Odysseus Group 1 grant awarded to the fourth author by the Flanders Science Foundation (FWO).

Ethical Approval: All procedures involving human participants were in accordance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards

Informed Consent: Informed consent was obtained from all participants

Availability of Data and Materials: The datasets analyzed during the current study are available from the corresponding author on reasonable request.

References

Alfonsson, S., Wallin, E., & Maathz, P. (2017). Factor structure and validity of the Depression, Anxiety and Stress Scale-21 in Swedish translation. *Journal of Psychiatric and Mental Health Nursing*, *24*, 154-162. doi: 10.1111/jpm.12363

Barnes-Holmes, D. (2018, July 10). The double edged sword of human language and cognition: Shall we be Olympians or fallen angels? [Blog post]. Retrieved from https://science.abainternational.org/the-double-edged-sword-of-human-language-and-cognition-shall-we-be-olympians-or-fallen-angels/rrehfeldtabainternational-org/

Barnes-Holmes, D., Barnes-Holmes, Y., Luciano, C., & McEnteggart, C. (2017). From IRAP and REC model to a multi-dimensional multi-level framework for analyzing the dynamics of arbitrarily applicable relational responding. *Journal of Contextual Behavioral Science*, *6*(4), 473-483. doi: 10.1016/j.jcbs.2017.08.001

Barnes-Holmes, D., Finn, M., Barnes-Holmes, Y., & McEnteggart, C. (2018). Derived stimulus relations and their role in a behavior-analytic account of human language and cognition. *Perspectives on Behavioral Science (Special issue on Derived Relations)*, *41*(1), 155-173. doi: 10.1007/s40614-017-0124-7

Barnes-Holmes, Y., McEnteggart, C., & Barnes-Holmes, D. (In press). Recent conceptual and empirical advances in RFT: Implications for developing process-based assessments and interventions. In M.E. Levin, P. Twohig, & J. Krafft (Eds.). *Innovations in Acceptance and Commitment Therapy.*

Baruch, D. E., Kanter, J. W., Busch, A. M., Richardson, J. V., & Barnes-Holmes, D. (2007). The differential effect of instructions on dysphoric and nondysphoric persons. *The Psychological Record*, *57*, 543-554. doi: 10.1007/BF03395594

Bentall, R.P., Lowe, C.F., & Beasty, A. (1985). The role of verbal behavior in human learning: II. Developmental differences. *Journal of the Experimental Analysis of Behavior, 43,* 165-181. doi: 10/1901/jeab.1985.43-165

Catania, A.C., Shimoff, E., & Matthews, B.A. (1989). An experimental analysis of rule-governed behaviour. In S.C. Hayes (Ed.), *Rule-governed behaviour: Cognition, contingencies, and instructional control* (pp. 119-150). New York: Plenum

Chiesa, M. (1994). *Radical behaviorism: The philosophy and the science.* Boston, MA: Author's Cooperative.

Harte, C., Barnes-Holmes, D., Barnes-Holmes, Y., & McEnteggart, C. (2018). The impact of high versus low levels of derivation for mutually and combinatorially entailed relations on persistent rule-following. *Behavioural Processes, 157,* 36-46. doi: 10.1016/j.beproc.2018.08.005.

Harte, C., Barnes-Holmes, Y., Barnes-Holmes, D., & McEnteggart, C. (2017). Persistent rule-following in the face of reversed reinforcement contingencies: The differential impact of direct versus derived rules. *Behavior Modification, 41*(6), 743-763. doi: 10.1177/0145445517715871.

Harte, C. Barnes-Holmes, D., Barnes-Holmes, Y., & McEnteggart, C., Gys, J., & Hasler, C. (2020). Exploring the potential impact of relational coherence on persistent rule-following: The first study. *Learning and Behavior.* doi: 10.3758/s13420-019-00399-0

Hayes, S.C. (1989). *Rule-governed behavior: Cognition, contingencies, and instructional control*. New York: Plenum

Hayes, S.C. (1993). Rule governance: Basic behavioural research and applied applications. *Current Directions in Psychological Science, 2,* 193-197.

Hayes, S. C., Barnes-Holmes, D, & Roche, B. (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition.* New York: Plenum.

Hayes, S. C., Strosahl, K., & Wilson, K.G. (1999). *Acceptance and Commitment Therapy: An experiential approach to behavior change.* New York: Guilford Press.

Hughes, S. & Barnes-Holmes, D. (2016). Relational Frame Theory: The basic account. In R. D. Zettle, S. C. Hayes, D. Barnes-Holmes, & A. Biglan (Eds.), *The Wiley handbook of contextual behavioral science* (pp. 129-178). West Sussex: John Wiley & Sons, Ltd.

Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the Depression Anxiety Stress Scales* (2nd ed.). Sydney: The Psychology Foundation of Australia.

O'Hora, D., Barnes-Holmes, D., Roche, B., & Smeets, P. M. (2004). Derived relational networks and control by novel instructions: A possible model of generative verbal responding. *The Psychological Record, 54,* 437-460. doi: 10.1007/BF03395484

O'Hora, D., Barnes-Holmes, D., & Stewart, I. (2014). Antecedent and consequential control of derived instruction-following. *Journal of the Experimental Analysis of Behavior, 102* (1), 66-85. doi: 10.1002/jeab.95

O'Toole, C. & Barnes-Holmes, D. (2009). Three chromometric indices of relational responding as predictors of performance on a brief intelligence test: The importance of relational flexibility. *The Psychological Record, 59,* 119-132. doi: 10.1007/BF03395652

Rosenfarb, I.S., Burker, E.J., Morris, S.A., & Cush, D.T. (1993). Effects of changing contingencies on the behavior of depressed and nondepressed individuals. *Journal of Abnormal Psychology, 102*(4), 642-646. doi: 10.1037/0021-843X.102.4.642

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech, Language, and Hearing Research, 14,* 5-13. doi: 10.1044/jshr.1401.05

Sidman, M. (1994). *Equivalence relations and behaviour: A research story*. Boston, MA: Authors Cooperative.

Sidman, M. & Tailby, W. (1982). Conditional discrimination vs. matching to sample: an expansion of testing the paradigm. *Journal of the Experimental Analysis of Behavior*, *37*(1), 5-22. doi: 10.1901/jeab.1982.37-5

Skinner, B.F. (1966). An operant analysis of problem solving. In B. Keinmuntz (Eds.), *Problem-solving: Research, method, and therapy* (pp. 225-257). New York: Wiley

Steele, D.L. & Hayes, S.C. (1991). Stimulus equivalence and arbitrarily applicable relational responding. *Journal of the Experimental Analysis of Behavior, 56*, 519-555. doi: 10.1901/jeab.1991.56-519

Zettle, R.D. & Hayes, S.C. (1982). Rule-governed behavior: A potential theoretical framework for cognitive-behavior therapy. In P.C. Kendall (Ed.), *Advances in cognitive-behavioral research and therapy* (Vol. 1: pp. 73-118). New York: Academ

Table 1. An illustration of the relational networks trained per experimental group.

	A	=	B	=	C	=	D	=	E	=	F	(= D)
High Coherence	LEAST LIKE	=	ZID	=	KROS	=	VEK	=	JUM	=	POM	= VEK
Low Coherence	LEAST LIKE	=	ZID	=	KROS	=	VEK	=	JUM	=	POM	≠ VEK
Control	FASTER/ SLOWER THAN	=	ZID	=	KROS	=	VEK	=	JUM	=	POM	= VEK

Fig 1

KROS is the same as ZID. ZID is the same as LEAST LIKE. **What does KROS mean?** "MOST LIKE" "SAME" "LEAST LIKE"	KROS is the same as VEK. VEK is the same as JUM. JUM is the same as POM. **Is VEK the same as POM?** "YES" "NO"

Fig 2

SAM is younger than TOM. TOM is younger than PAT. **Which is the oldest one?** "PAT" "PAT" "PAT"	A is stronger than B. B is stronger than C. **Which is the strongest one?** "A" "B" "C"

Fig 3

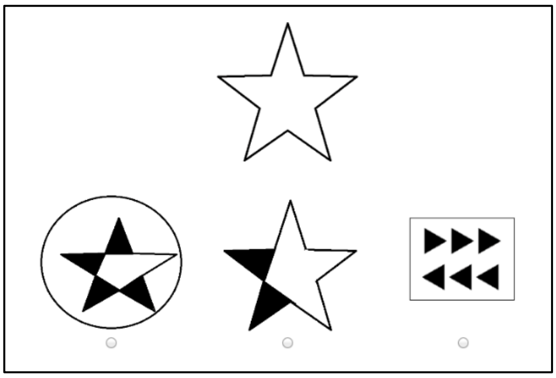
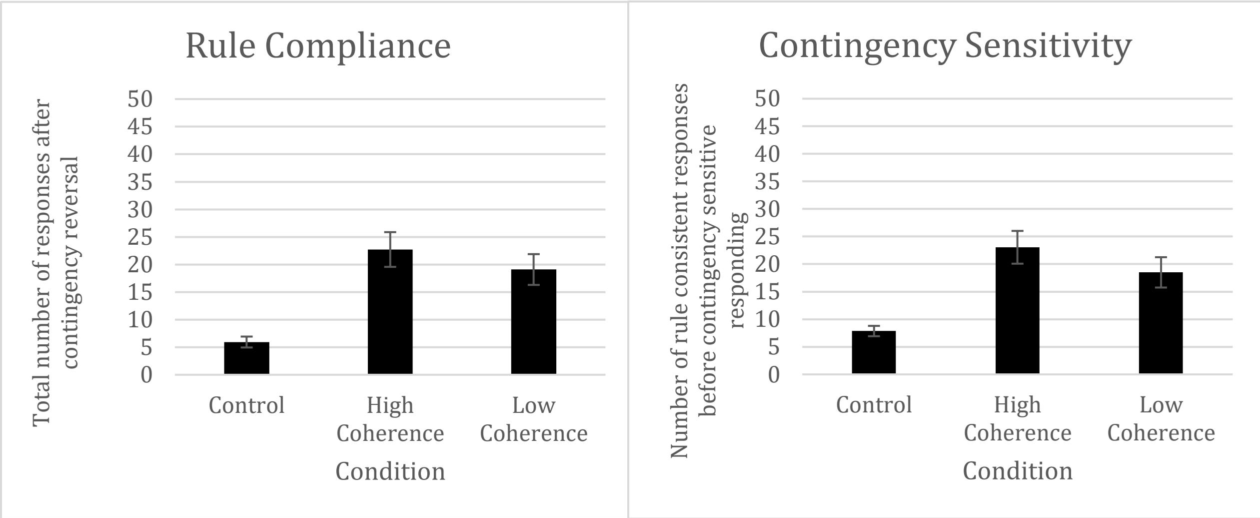
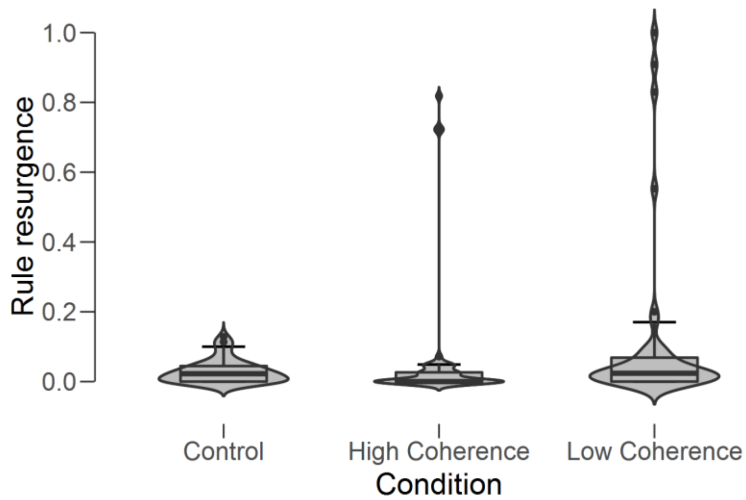


Fig 4



Rule Compliance
Total number of responses after contingency reversal
50
45
40
35
30
25
20
15
10
5
0
Control
High Coherence
Low Coherence
Condition
Contingency Sensitivity
Number of rule consistent responses before contingency sensitive responding
50
45
40
35
30
25
20
15
10
5
0
Control
High Coherence
Low Coherence
Condition

Fig 5



Rule resurgence
1.0
0.8
0.6
0.4
0.2
0.0
Control
High Coherence
Low Coherence
Condition